DOCUMENT RESUME

ED 091 821 EA 006 175

AUTHOR	House, Ernest; And Others
TITLE	An Assessment of the Michigan Accountability System.
SPONS AGENCY	Michigan Education Association, East Lansing.; National Education Association, Washington, D.C.
PUB DATE	Mar 74
NOTE	32p.
EDRS PRICE	MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS	*Citizen Participation; *Educational Accountability; Educational Assessment; *Educational Objectives; Program Evaluation; *State Departments of Education; *State Programs
IDENTIFIERS	*Michigan

ABSTRACT

Accountability currently is an important concept throughout education. Michigan has assumed leadership among the States in exploring and applying accountability procedures. The purpose of this report is to examine the quality and implications of that leadership. Specifically, the authors' purpose is to assess the Michigan Accountability System with respect to the educational soundness and utility for Michigan and with particular emphasis on the assessment component. The report describes the system and offers recommendations for improvement. (Author/JF)

ED 091821

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

An Assessment of

The Michigan Accountability System

by:

Dr. Ernest House
University of Illinois

Dr. Wendell Rivers
University of Missouri at St. Louis

Dr. Daniel Stufflebeam
Western Michigan University

under a contract with:

The Michigan Education Association
The National Education Association

March, 1974

EA 006 175

ACKNOWLEDGEMENTS

We are grateful to the many persons and groups who contributed to this study. Both oral and written testimony were supplied by many teachers and administrators and by representatives of civic and professional associations. Michigan Department of Education officials were cooperative, open, and professional in testifying and in supplying many pertinent printed documents. The Michigan Education Association and the National Education Association representatives supplied excellent logistical support, and their professionalism was of the highest order in keeping their input of information to the committee on the same level as existed for all other persons and groups. We were greatly assisted by all. We think this testifies strongly to the balance of professional concern and cooperation that exists throughout Michigan regarding the continuing development of the state accountability system.

INTRODUCTION

Accountability currently is an important concept throughout education. Clearly Michigan has assumed leadership among the states in exploring and applying accountability procedures. The purpose of this report is to examine the quality and implications of that leadership. Specifically our purpose is to assess the educational soundness and utility for Michigan of the Michigan Accountability System with particular emphasis on the assessment component.

Our study was supported by the Michigan Education Association and the National Education Association. While these organizations secured our services, we have retained complete independence in writing, editing, and releasing this report. (A copy of our working agreement with MEA/NEA is appended to this report.) This document thus represents our views and not necessarily those of any other party.

In preparing this report, we gathered, received, and discussed an extensive amount of information about the Michigan Accountability System. We obtained and studied a large number of published documents. We heard more than thirty hours of direct testimony presented by persons representing all levels of Michigan's educational system. We received and studied specially prepared written testimony, and jointly prepared this report.

Our efforts were unified through a common view of the importance of educational accountability. We believe that accountability should be practiced at all levels of education. We believe it should serve both to improve and to prove the quality of education. We believe that different conceptions of educational accountability need to be tested under

field conditions, and that experimental efforts in accountability should be critically examined prior to widespread implementation. Hence we welcomed this charge to participate, through the role of critic, in an effort to advance the practice of educational accountability in Michigan.

We hope this report will be used by citizens, state and local board members, legislators, and educators throughout Michigan to improve the state's accountability system. Our purpose will have been served if this report stimulates thoughtful discussion about the strengths and weaknesses of the present state accountability system. More than this, we hope that our analysis will point the way to improvements in the Michigan Accountability System.

STATE LEVEL LEADERSHIP

Leadership and innovative activities often are marked by controversy. The Michigan Accountability System is no exception. Personnel in the Department of Education clearly have exerted courageous leadership in shaping the Michigan Accountability System, and their activities have been the center of much controversy in the state.

Clearly, Michigan is fortunate in the quality of staff in its Department of Education. Its staff possesses significant experience and expertise in educational administration, testing, and research methodology. Moreover, the staff has made extensive use of experts outside the Department. Any faults in the conception and implementation of the state accountability system are not attributable to the qualifications of the persons who developed the program; nor do we find any fault with their motives.

However, educational accountability is not easy to work with. Educational researchers have not produced tested standards and procedures for state accountability systems. But many states' legislatures have mandated that systems of accountability be implemented on a crash basis. In Michigan, the state staff has responded not by inertia but action. Consequently, they have often implemented new accountability procedures without taking the needed time for conceptualization, development, and testing. Their motto seems to have been "if in doubt, go ahead." Not surprisingly the state staff has been severely criticized from many sectors for having moved too rapidly without an adequate rationale.

While we salute the competence, motives, energy, and innovative spirit of the Michigan Department of Education staff, we believe they have made serious errors in their past efforts to implement statewide accountability. We believe gentleness in criticism is not a kindness. Therefore, the remainder of this report is often pointedly critical. We hope that our analysis will serve Michigan citizens through stimulating needed changes in state level accountability activities.

THE SIX-STEP ACCOUNTABILITY MODEL

Michigan's six-step accountability model has a number of appealing features:

1. Involving persons from throughout the state in defining common goals is a useful way of focusing communication about educational accountability.

2. Translating common goals into objectives potentially provides a broad base of important variables for assessing needs in Michigan's schools.

3. Assessing needs in relation to objectives derived from the common goals should provide information to state and local level decision makers to help them determine priorities for a variety of needed change efforts.

4. Testing alternative delivery systems should assist the state to develop a research base for assisting schools to adopt innovative strategies that will serve high priority needs.

5. Fostering the development of local evaluation capability should assist the schools to assess local needs; to design, implement, and assess their innovative efforts; and to evaluate their personnel on fair bases.

6. Using feedback from the accountability system to guide state and local educational policy should assist school districts and the state department of education to fulfill their leadership roles in education.

Our reservations about this model are not in its rhetoric but in its implementation. While the state has made some desirable progress in implementing the model, we believe a number of activities have not been consistent with the intent of the model and have in fact been counter-productive. In the remainder of this section we present our views concerning the state department's efforts to implement each of the six steps in the state accountability model.

The Common Goals

The common goals are broad and generally non-controversial. As one Michigan educator testified, the common goals are like the Seven

Cardinal Principles of Education.

However, the common goals are unclear, and there seems to be no on-going review of the goals and provision for updating them. The goals contain redundancies, and they are not definitive. While some of the goals are defined through examples, generally the examples are broad and give little direction for developing specific objectives. These criticisms are not crucial since they pertain to the technical as against the philosophical qualities of the goals.

However, the common goals are the basis for specific objectives and needs assessment efforts at both the state and local levels. Because of their functional importance in the Michigan Accountability System, the common goals should be made as clear as possible; and periodically they should be reviewed and updated.

The Objectives

The Department of Education has attempted to translate the common goals into performance objectives. Their activities so far have been in the cognitive and not the affective areas, which is a significant limitation but not necessarily a fault. To its credit, the state staff has sought to secure wide involvement of citizens in Michigan for the development of objectives. However, the state staff has made exaggerated and untenable claims for the results of their objectives-development work. It is important that these faulty claims be exposed since they could lead to misinterpretation and misuse of the objectives and the objectives-referenced tests.

This is true especially in the areas of reading and mathematics. In these areas state officials have selected 23 and 35 objectives, respectively, for the fourth grade; and 23 and 45 objectives, respectively, for the seventh grade. In presenting testimony to our panel the state department representative indicated:

> "The performance objectives in reading and mathematics and in the other areas should be viewed as a consensus among educators at all levels of the educational system and in all regions of the state as to the minimum behaviors that students should be able to demonstrate at selected levels of the educational continuum."

Unfortunately, neither the procedures by which the objectives were derived, judgments of the objectives themselves, nor the test results support these claims. In the same testimony they indicated that practical considerations partially dictated how many and which objectives were finally chosen. For example, they testified that "a decision was made that if fewer than five items for an objective appeared to be sound, then the objective was eliminated." Also the objectives were reduced to the number that could be tested for within a few hours. Clearly, the objectives so far developed do not represent a consensus of educators and the objectives are not minimal.

Only a relatively few persons have been involved in developing and choosing the objectives. Sampling procedures have not guaranteed that these persons are representative of the large population of persons who are concerned with education in Michigan. Moreover, testimony presented to our panel indicates that the objectives that were chosen do not represent consensual choices of even the small group of persons who were involved in the development of objectives. It is certain that a consensus among Michigan educators has not been reached concerning minimal performance objectives; and it is doubtful that such a consensus could be achieved.

Likewise the claim that the objectives are minimal is unfounded, and there is considerable reason to believe the objectives are not minimal. The following quotation appears in the California Test Bureau's Technical

Manual concerning their work in assisting the Michigan Department of Education (CTB/McGraw-Hill[1] p.31) to develop the new objectives-referenced tests:

> "For a test with objectives pertaining to instruction at the level at which the students were tested, we would expect to find at least 50 cases in the $f_1 + f_2$ cell of Category 3 (those who had received instruction pertaining to the objective). Unfortunately this is not true of the Michigan data. Even testing the grades below the level of intended use did not provide a sufficient number of Category 3 cases to allow sensitivity to instruction data to be used to validate these tests. ...it would appear that a controlled study of the objectives would be highly desirable."

Also Dr. Frank B. Womer has reviewed assessment results based on the administrations this fall of the new objectives-referenced tests. He noted that "...if one accepts at face value the information now available in assessment publications and from public statements about Michigan assessment," one may conclude that "...not a single school district in this state is meeting minimal objectives in reading and mathematics."[2] He goes on to say that "...such a conclusion is unwarranted" and urges that the prejudgment that all of the assessment objectives are truly minimal should not be accepted without careful, personal evaluation. Still further questions about the extent to which the new objectives are minimal were raised by representatives of the Bloomfield Hills School District. Although their students traditionally score higher on tests than students in most other districts in Michigan, for some of the objectives in the new objectives-referenced tests less than 50 percent, and in one case less than 30 percent, of their students passed some of the objectives. Thus, the objectives selected by the Michigan Department of Education are not minimal.

We recommend that the Michigan Department of Education staff modify

their claims that the selected objectives are minimal and represent a statewide consensus. Instead they should urge caution in the use of and interpretation of test results related to the objectives. Further, we hope the state staff will abandon its recently announced plans to publish a book of objectives for parents. The reported hope of state department officials that the book will provide a handy reference for parents to check up on their children's progress in school seems unfounded. Based on the state's performance so far in trying to choose minimal performance standards that represent a statewide consensus, it is more likely that such a book will lead parents to develop faulty assumptions concerning what their children are being taught and unrealistic expectations concerning what their children should be achieving at given grade levels. Worse, the book could lead to a state controlled, monolithic curriculum. The prospects for misuse and misinterpretations of the state objectives are not unlikely and the possible consequences are not trivial. Educators throughout Michigan should require that state leadership personnel act thoughtfully and responsibly in describing and using the state objectives.

The Assessment Component

The most serious breakdown in implementing the Michigan Accountability Model is in the assessment component. The model's promise of providing on-going needs assessment in relation to the full scope of the common goals has not been pursued. Instead, attention has been limited mainly to reading and arithmetic at two grade levels. To no constructive purpose schools have been ranked on norm-referenced tests. Objectives-referenced

tests have recently been put into full-scale use before being validated. All pupils are being tested when there is no compelling reason for this. There is a present danger that the weakness of the assessment component may undermine the total structure of the accountability model. Overall, we see so many serious problems in the state's implementation of the assessment component that a later section is devoted entirely to the assessment activities.

Analysis of Delivery Systems

Step 4 of the state accountability model points the way to school-based innovation and experimentation. The legislature and state department have fostered much activity in this area of the model. We find some of this activity to be appropriate and good, but we think the part that relates to the Chapter 3 Program has serious flaws.

On the positive side, we support the state-sponsored research and development work being conducted to identify and analyze alternative educational practices. It is also commendable that funds are being concentrated on the basic skills problems of disadvantaged children. Results from such improvement-oriented activities should be of use to those educators who have the difficult assignment of improving educational experiences for poor achievers.

However, we have serious reservations about the implementation of the Chapter 3 Program. We believe this program is potentially harmful to education in Michigan, in tying money to test scores. School districts are told that they will be rewarded if their poor achievers attain state minimal standards in reading and mathematics. Since the districts are not given the funds until and unless the students meet the state standards,

the implication is that educators are not doing their work well, but that they can and will improve their performance under the promise of financial rewards. We think this is a gross misinterpretation of the problems in educating disadvantaged children; and we think the implications in the Chapter 3 Program about the professionalism of Michigan educators are wrong and demeaning.

There are also measurement and statistical problems in the Chapter 3 Program. It seems certain that financial rewards to school districts are often given and withheld based on measurement error. This is especially likely because of the use of gain scores and because the students being tested are at the bottom of the state distribution. We also find dubious the claims that Michigan's Chapter 3 Program has produced real gains in achievement for disadvantaged children. This claim could be established only through the conduct of a rigorous field experiment.

Overall, we support the emphasis being given to improving education for disadvantaged children. But we recommend that the state abandon its practice of rewarding school districts for good test performance of their disadvantaged students. In a later section we deal in more detail with the problems in the Chapter 3 Program.

Developing Local Evaluation Capabilities

School districts need assistance in developing systems for evaluating their programs and personnel. The Michigan Accountability Model is laudable in promising assistance in these areas. We encourage the state department to greatly expand their activities in implementing step 5 of the Accountability Model.

Use of Accountability Data for Decision Making

State and local level decision making are receiving little service from the state accountability system. Decisions to be served have not been clarified and there is no on-going procedure for determining state and local information requirements that should be served by the accountability system. Neither is there evidence that the Governor, legislators, or state board members have used Michigan assessment information to shape educational policy for the state. Particularly there is no reason to believe that testing all pupils at specified grade levels on all test items in reading and mathematics provided vital information to any group in the state. Considering the great cost of testing all pupils (especially to the pupils and their teachers) we urge that this practice be abandoned until there is clear cause for it.

In regard to clarifying the purposes of the state accountability system, and especially the assessment component, we believe that Dr. Frank Womer is providing constructive service to the state. His analysis of possible goals for state assessment provides a valuable focus for clarifying what services should be provided to what groups. We think the state should seriously consider his suggestions that sampling as opposed to every-pupil testing may be sufficient to meet the purposes of state accountability.

Overall, the Michigan Accountability Model is an appealing conceptual scheme, but has been poorly implemented. The most serious weaknesses are in the assessment and Chapter 3 activities. We consider the assessment component in the next section.

STATE ASSESSMENT

The Michigan Educational Assessment Program is based upon the use of

objectives-referenced instruments which, theoretically, provide information concerning student performance in relation to a specified set of learning objectives. Two important factors appear to have underscored the need for an alternative to the previously used, norm-referenced tests. First, testimony received by the panel indicated that unfair comparisons of student performance between districts occurred as a result of the manner in which the results of these instruments were reported; and that the test, in fact, contained content which discriminated against certain student groups within the system. This was particularly true for minority group youngsters.

Second, in addition to the many technical deficiencies inherent in norm-referenced instruments, there appears to have existed a general consensus among responsible state and district personnel that these instruments did not measure that which Michigan teachers were teaching in the schools. The change from norm-referenced instruments to the present use of objectives-referenced tests has met with general approval throughout the Michigan educational community.

The Michigan Department of Education is to be congratulated for its attempt to develop a new type of assessment approach in the form of objectives-referenced tests. This step represents a bold, new innovation and a worthy experiment. However, much of the controversy surrounding the assessment program results as much from the way the program is being implemented as well as from its substance.

A number of important critical concerns regarding the assessment component emerged within the testimony received by the panel. These

concerns related to the scope and general appropriateness of the tests and the subsequent use of the performance data from their administration on a statewide basis. This section of our report focuses on these expressed concerns.

The Tests

The objectives-referenced tests have admirable reliability. That is, KR-20 and point biserial correlations are acceptable for most objectives and items.[3] In other words, the five items used to measure each objective seem to be measuring the same thing, in a consistent manner.

What the items are measuring is the question of validity. This is a problematic area as far as these tests are concerned. While there is reasonable fit between the items and the objectives, the tests can only be as good as the objectives themselves. Unfortunately, it appears as if the test developers did not establish the validity of the objectives before putting the tests into use. An attempt to develop a special statistic, "The Sensitivity Index," as a measure of validity did not meet with success. Hence, we must conclude that the reliability of the tests is good but that their validity is questionable and requires further examination.

Of equal cause for concern is the assumption inherent in the statewide approach to assessment that the same test instrument form can be used to assess learning in all of the fourth and seventh grade children in the state. Evidence is fast accumulating that certain children are penalized by assessment instruments which were designed to accommodate the language and experiential backgrounds of a majority of the population and which do not have equivalent forms which accommodate these factors in a sizeable minority population of children (Williams, 1972).[4]

This panel found no evidence that ethnic membership or economic levels were considered as important variables in the development of the present forms of the assessment instruments. In addition, no evidence was available that the items comprising the reading test had been field tested or validated for use with children whose sub-cultural language styles differ from those of the middle-class, ethnic majority Michigan children. We strongly recommend that the presently used reading assessment devices be thoroughly tested and validated for use with minority and low-socio-economic children before considering their performance as indicative of their true potential. We suggest that the diagnostic value of these instruments will be greatly enhanced if such a validation is immediately instituted.

Involvement of Teachers in Developing Tests and Objectives

Among the thirteen panel members who helped develop and review the objectives, four were teachers. Also, some professional associations were sent the objectives and items to review. However, even those teachers who participated felt their involvement was rather cursory and too much a matter of form. Members of the reading panel have described their involvement as making no significant difference (MASB, Jan. 1972).[5] Strong opposition and resentment expressed by many teachers and some teacher groups indicates a feeling on their part that they were not significantly involved.

Cost of Every Pupil Testing

Statewide testing as presently executed also raises the question of the feasibility of every pupil testing. This practice appears to be of dubious value when the cost of such an undertaking is compared with the

resulting benefits to local level personnel. Evidence gathered by this panel, from the verbal testimony presented, indicates that plans are presently being formulated which will lead to the expansion of the assessment program to other grade levels and other subject areas. If every pupil is tested at several grade levels, on all of the variables implied by the common goals, then the eventual costs could be millions of dollars. While administration costs to the SDE might run as little as $100,000 to $200,000 per subject area per grade level, the administration and educational time of the local schools must be calculated in the time costs. Four or five hours per child per subject area represents a heavy investment indeed. Unfortunately, we were not able to obtain cost projections from SDE.

Other costs include the loss of instructional time and excessive duress on the students. Some administrators and teachers feel that more testing is the last thing they need. Some report that much of September is consumed in preparing for and administering both state and local test batteries. Some teachers also report extreme frustration on the part of students who must take the tests, particularly those who have a difficult time with the highly verbal items.

The local, and hence overall, costs could be reduced by a matrix sampling plan which requires that each student tested take only a few items. This would provide accurate statewide data, but not data on local schools or individual students. In the long run, a matrix sampling plan will be the only one feasible from a cost and time standpoint. The cost and time required for every pupil testing for the whole state would be horrendous. In actuality, those most closely associated with the testing program in the Department of Education prefer matrix sampling.

We strongly believe that strict adherence to a statewide testing model is detrimental to the legitimacy of the Michigan Assessment program. We feel that it will result in useless expenditures of monies and manpower, in addition to producing unwarranted disruptions of the educational programs within a great number of schools.

Domain Limitations of Present Instruments

The presently utilized state assessment instruments are limited in scope. It is understandable that reading and mathematics are vital skills and must be mastered if a child is to achieve learning benefits from the broad spectrum of academic experiences to which he will be exposed. However, to utilize these skill areas as the singular criterion of scholastic progress, through the use of narrowly structured objectives is grossly unfair to the child and similarly so to the teacher of the child. Therefore, a mandated set of rigidly applied performance objectives violates the existence of a broad spectrum of human learning styles, strengths, and teaching techniques. In any case, testing all pupils in two grade levels in the areas of reading and mathematics does not respond to the supposed purpose of broad based needs assessment. It seems that this purpose has been abandoned in favor of routine testing which has been done for decades, and is relatively easy to do.

Need for Locally Developed Objectives

According to testimony, there is some threat to the local curriculum from the state tests. Some art teachers have been told to teach math and reading. Some principals have been told to raise the scores of their schools or lose their jobs. Some schools have taught the items on the test

in preparation to taking them. How much the curriculum has been perverted or how much cheating there really is is difficult to say at this point.

It is important that local autonomy be preserved. As some witnesses pointed out, true reform must occur at the school level. While the state can stimulate and help, education occurs in the school at the local level with particular children. Instruction must be fitted to the personal needs of these children and cannot be determined "a priori." We strongly recommend the importance of locally developed objectives.

Publication of Test Scores

It is our understanding that, to date, listing of district test performances have been published. Few incidents have excited as much controversy as the publication of district test scores with districts ranked from top to bottom. Such rankings invariably reveal the socio-economic ordering of the state by school district. It is our opinion that the continuation of such practices could lead to highly detrimental consequences for the Michigan educational system as a whole. Probably the most damaging effect of this practice would be the misinterpretation of this data by those whose motivations are based upon factors other than of the improvement of education in the state of Michigan. The scores are largely misinterpreted by the public. Low test scores are taken as a sign of a poor educational system. Part of one district wanted to secede from another on the basis of the fall results.

Ample examples of misinterpretations abound in the newspapers. For example, two Detroit newspapers reported that the bulk of Detroit's students were reading at the second percentile because the district as a whole ranked at the second percentile among districts of the state--a

gross misinterpretation. Unfortunately, the public interprets test scores as literal indicators of what is taught in school, which they are not.

The rationale for publicizing the test scores is that it will excite the parents to do something about their school. Yet, districts such as Detroit have been publishing test scores for years and are still at the bottom of the list.

Teacher Evaluation

Possibly one of the most unfortunate potentialities of the Michigan Assessment Program would be the use of student test scores as the major criterion for the evaluation of classroom teachers. Many of the factors which influence learning are not controlled by teachers. This is particularly true for such factors as the background experiences of the student, his emotional and physical readiness for school, the cognitive and affective skills which he brings from his particular family milieu; and numerous other personal and school-related factors. The present state of the art in psychometrics and test development does not allow tests, objectives-referenced or not, to adequately assess and document the impact of these factors on a child's performance.

Needless to say, tests on which the validity is questionable should not be used to evaluate teachers. What if the tests were better? Even so, tests should not be used to evaluate teachers. This is not to say that teachers should not be evaluated. We believe they should. However, test results are not the way to do it. Test results are not good measures of what is taught in school, strange as it may seem. They are good indicators of socio-economic class and other variables. But, unless one teaches the tests themselves, they are not very sensitive to school learning.

We have not uncovered any instances in which teachers have been fired as a result of statewide tests. There is, however, considerable anxiety, distrust and fear among teachers and many feel personally threatened. If the SDE puts books of objectives into parents' hands, this anxiety might be expected to increase tremendously. Parents will be asked to check whether their child has learned those objectives which have not been validated for that grade level or school; consequently, considerable conflict may be expected to develop between teachers and parents.

It is hoped that responsible state department administrators will give immediate attention to the presence of those fears described above and also to the negative potentialities of this practice. Immediate steps should be taken to develop safeguards against the emergence of such a detrimental and demoralizing practice.

Value of the Assessment Program for Various Audiences

Perhaps the most unexpected finding is that the assessment program has little apparent value for any major group. Most seem to regard it as unimportant or a necessary evil. Many teachers do seem to find the results more useful for instructional purposes than norm-referenced tests. Their enthusiasm for the usefulness of this device is considerably less than overwhelming. Administrators seem to feel that the last thing they need is another test. They already administer their own local tests.

Districts like Detroit have their own tests and also have been making test scores public for years--with no dramatic jump in achievement. High achieving districts have little need for a test that purports to measure "minimal skills." While one of the strongest features of the tests is

its forms for reporting results, response in the field is not particularly positive.

Higher level decision makers are also unsure about how the results will be or can be used; however, they feel that the publication of the data may incite parents to do something about the schools. The state board feels, somehow, the program should function as a positive incentive, but how this will happen is not clear.

One high level state official said the hope is that the test results will force a school to do better because of pressure from the parents. Both the publication of district test results and the expected dissemination of "objectives" books to parents appear to follow this logic. We oppose this tactic. In the long run such assaults upon the teachers and schools are likely to force educators into a very defensive posture and cause them to concentrate on public relations rather than on substantive innovation. The long-range effects of such "forcing" policies are likely to be counter-productive to better schools.

What can an assessment program do? We can imagine situations where data might be useful for state decisions, as for example, determining what types of state programs to mount. This statewide information could be provided most effectively and cheaply by a matrix sampling plan. We strongly recommend a switch to a matrix sampling assessment program and the curtailment of every pupil testing.

There are also some schools and individual teachers who see benefits in the total implementation of the accountability-assessment approach. The SDE should provide strong assistance and encouragement to these schools, help which they now need and are not getting. An important part of the

SDE function might be to help those schools that want such an approach, but not to force schools that do not. While the SDE has the authority to demand that the schools do something about their basic problems, and even to provide evidence that they are doing something, the SDE certainly does not have the knowledge or expertise to solve all the schools various problems through some statewide solution.

There was very little testimony, outside of that given by the state department personnel, which supported the Michigan Assessment Program. The panel was provided with information indicating that this posture of non-support is widespread throughout the Michigan education community. As unfortunate as this situation is, satisfaction must be taken in the fact that this non-support is a reflection of the concern that Michigan educators have for the welfare of all of the children of Michigan, particularly in the face of the potential massive negative impact of a widespread testing program. We share this concern.

The assessment component needs to be reconceptualized and reorganized and its purpose clarified. An overall effort should be expended to assess the needs of the audiences to be served, and these audiences should be given a part in the determination of the structure and function of the total program.

CHAPTER 3 PROGRAM

Chapter 3 of the State School Aid Act provides special funds to school districts with heavy concentrations of low achievers. An extra $200 in student aid per child identified as being educationally deficient--essentially being in the bottom 15% of the pupils on achievement tests--is provided

to these districts. This is in addition to federal compensatory aid and is to be highly commended as an action to improve the plight of the disadvantaged. Such fine intents motivate much of the SDE activity. Unfortunately, the second part of Chapter 3 is not so laudatory.

A school district must show at least three-quarters of a year gain in individual achievement scores or be penalized part of the money. The district must produce a different "delivery system." After the delivery system has been implemented students are tested to determine whether they have met state standards. If they haven't, their district does not receive its allocation of Chapter 3 funds.

Paying on the basis of test scores is whimsical at best. Contrary to public opinion, standardized achievement tests are not good measures of what is taught in school. In addition, many other factors outside school influence them. Even on highly reliable tests, individual gain scores have only a fraction of the reliability of the test itself. This means that individual gain scores can and do regularly fluctuate wildly for no apparent reason by as much as a full grade equivalent unit. Money may be awarded not on true gain but on "test error." For tests that are not fully developed like the assessment tests, the problem could be greater. The 75% achievement standard is purely arbitrary and one for which no one claims responsibility.

In order to prove that the test gains did result from instruction, it would be necessary to employ a rigorous experimental design involving randomized control and experimental groups. This is very difficult to do. Even if the tests were completely valid and reliable, it would not be possible to attribute achievement gains to the school or teacher.

We hope that state awards of compensatory funds be continued and that the search for new delivery systems will also be continued. Unfortunately, the alternative delivery systems have consisted of performance contracting which incorporates the same payment-by-test-results errors that Chapter 3 now encompasses. Because of its known and proven deficiencies, performance contracting has been abandoned in most of the country. We hope Michigan will follow suit. We repeat: educational deficiencies will not be remedied by taking money away.

It should be noted that most witnesses were opposed to the tying of test scores to money. It is the point of greatest consensus among those who testified. As Chapter 3 now stands, it allows school districts to be punished if the children who qualify under Chapter 3 do not perform on the tests at levels dictated by the state. This is an untenable practice.

SUMMARY

One concerned state official has said: "Sometimes I wake up at night worried about the direction we're going. It's like a giant snowball rolling downhill." This statement captures the feeling of the panel toward Michigan accountability and assessment. Driven by intense political pressure from the Legislature and the Governor's office, the SDE leadership has launched a bold and well-intentioned effort to reform educational practice. Unfortunately, it was not thoroughly thought through and it has gathered a momentum somewhat lacking in thoughtful control.

The Michigan accountability model itself has many good features. It has stimulated a public discussion of the goals of education and provided direction for state accountability efforts. It has involved educators throughout the state in efforts to develop objectives and it has resulted

in pilot forms of objectives-referenced tests that some teachers have found useful. Overall the state's accountability work has created an aura of innovation and change.

However, in implementation of the accountability system, many activities have been inconsistent with the model's intent and even counterproductive. The common goals have not been clarified; the objectives were developed by relatively few people and do not represent either a consensus or minimal objectives; and school districts nave been given little help in developing school-based evaluation systems. The most serious flaws are in the assessment component; it is too narrow in scope to serve as a state needs assessment, and it has been implemented on an every pupil basis without technical or utility justifications. Overall, there is no clear evidence that state and local decision making is being served by the accountability model.

The broad educational goals were tested for in a few cognitive areas--on basic skills at a few grade levels. While the objectives developed were not bad, the claim that they represented a consensual agreement for Michigan of "minimal objectives" is false. Nor were they adequately field-tested or validated. The items used to measure those objectives had good reliability but questionable and undemonstrated validity. Development of both objectives and test items was so hurried that only minimal involvement of teachers was possible.

In an effort to avoid the criticism raised by the previously used norm-referenced tests, the new unvalidated objectives-referenced tests were put into immediate use all over Michigan. The results testify to the crudity of the tests. At each level each development group did the

best job possible under the circumstances, but the premature use did not allow for proper development. The first set of results has dismayed many teachers, administrators, and educational groups.

These events have fueled the previous mistrust of the SDE which dates back to the publication of district test comparisons. Added to the fire has been the tying of test-score results to funding, a whimsical procedure at best and possibly quite a harmful one. Although these measures were initiated with a view to focusing attention on the educationally deficient student, an admirable motive, the effects are more symbolic than real. The millions of dollars might be better spent in a number of other ways.

The accountability and assessment systems also risk substituting state objectives and curricula for local ones, not a reasonable educational trade-off in our opinion. Teachers should try to teach to the child, not to state tests. We know that none of these ill side effects is the intent of the SDE or the state board, yet it is common for side effects of treatments to outweigh the benefits, how ever noble the motives of the sponsor or how ever real the problem.

In this respect we have made a number of specific suggestions that we think would improve the Michigan assessment though we confess that, beyond the small group of developers themselves, support for the assessment program is not strong or widespread. In fact, it is difficult to see exactly what decision situations the assessment will serve. We have suggested slower development, matrix sampling, putting the assessment on a voluntary rather than mandatory basis, and placing more emphasis on the development of school-based evaluation. Eventually this should improve

the assessment's quality and utility.

Involving more educators and providing assistance to those who want to implement the full accountability model is strongly recommended. This would put the assessment in more of a market situation in which it must demonstrate the utility of its services rather than mandate them. We strongly recommend against the absurd practice of tying money to gains in achievement scores.

Again we laud the SDE on its bold and innovative leadership in attacking some of the major educational problems of our time. We are sincere in believing the Michigan staff to be as competent and highly motivated as that of any state education agency in the country. In particular, state education agencies often suffer from a failure of imagination and nerve. We are happy to say that is not the case here. We do hope that the SDE will realize that such admirable, aggressive behavior in attacking complex problems often results in errors. One of the ways of catching such mistakes and correcting them is by listening carefully to what others have to say.

REFERENCES

1 1972-1973 Test Development and Experimental Tryout, Technical Report for the Michigan Department of Education, April, 1973.

2 Frank B. Womer "Interpretation of Michigan Educational Assessment Results." Bureau of School Services, U. of Michigan, January, 1974.

3 CTB Report 1972-1973 Test Development and Experimental Tryout, Technical Report for the Michigan Department of Education, April, 1973.

4 Williams, Robert L. "Causes and Misuses in Testing Black Children." The Counseling Psychologist, 1971, 2 (1), 62-73.

5 A. Gagnon, Michael J. Homes, and Lawrence F. Reed. "A Sub-Committee Report." Michigan Association of School Boards Journal, Jan. 1972.

APPENDIX

(Memorandum of Agreement Between Blue Ribbon Panel and MEA/NEA)

1. Charge

The external evaluation panel consisting of Dr. Wendell Rivers, Dr. Ernest House, and Dr. Daniel Stufflebeam has been engaged by the Michigan Education Association and the National Education Association to evaluate the educational soundness and utility for Michigan of the Michigan Accountability Model with a particular focus on the assessment component.

2. Audiences (in priority order)

NEA/MEA.
Decision makers in Michigan's educational system (State Board of Education and State Department of Education).
The media (the public).
Consumers (parents, PTA, the public, etc.).
Technical persons (especially in the area of Educational Measurement).

3. Report/editing.

The panel will be solely in charge of developing and editing its final report. NEA/MEA may write and disseminate any separate statement (such as an endorsement, a rebuttal, a commentary, or a descriptive piece). It is understood that the panel's report is to be as short and direct as possible and to be designed to communicate with the audiences designated for the report.

4. Dissemination

The external panel has the right to release its report to any members of the target audiences or other persons following the completion of the report. The panel's release of the report will imply no MEA/NEA endorsement. Further MEA/NEA may choose to endorse or not endorse the report depending on their judgment of the quality and appropriateness of the report. Should MEA/NEA decide to disseminate their own document describing the report their document will be identified as their own and not that of the committee. Only the committee's final report as edited by the committee will be distributed with the names of the committee on it.

5. Format of the Report

The following items were identified as desirable ingredients for the panel's final report:
a. citation of the agreements between the review panel and NEA/MEA
b. presentation of the major findings
c. presentation of minority opinions, if any.

6. Questions to be Addressed in the Report

Specific questions to be addressed will include:

a. validity and reliability of criterion-referenced tests
b. use of tests to evaluate staff
c. merit of the objectives on which Michigan assessment is based
d. involvement of teachers in developing both objectives and tests
e. the panel's recommendations for change and further study
f. comments about the balance of the state effort and appropriateness of expanding the scope of assessment especially given cost factors associated with the projections for improving or expanding Michigan assessment
g. quality of planning in the Michigan Accountability Program
h. cost benefit projections for the program
i. value of Michigan assessment outcomes and reports for different levels of audiences in Michigan
j. problems of bias in the Michigan Accountability Program.

7. Resources (budget) to Support the Program

Sufficient resources will be made available by MEA/NEA to the external review panel to support eight days of work per panelist to work on the program, whatever secretarial support is needed in conducting the program and whatever materials and equipment requirements, for example, tape recorders, taping, etc., in the Lansing hearings. It is understood that if any of the panelists need to make long distance telephone calls to collect opinions about the program from people in Michigan that the panelists will be reimbursed for such expenses provided that an accurate and complete report is made of the purpose of the phone call and who was contacted.

8. Delivery Schedule

The panel is to deliver its final report on March 1 or as soon thereafter as is practicable.

9. Access to Data

It is understood that the Michigan Department of Education will make available to the panel any and all data and reports required by the panel to do the job. This, of course, is restricted to those data and reports that are now available to the Michigan Department of Education regarding Michigan accountability.

10. Procedures

Pursuant to the above conditions the external three-man panel will have control over the evaluation process that it must implement to responsibly respond to the charge to which it has agreed. In accordance with this position the panel has agreed to implement the following general process.

Private interviews and hearings will be conducted solely by the panel

with representatives of the Michigan Department of Education, representatives of NEA/MEA, representatives of selected groups (teachers, administrators, board members, and educational action groups). The panel will also review documents made available to it by NEA/MEA and the Michigan Department of Education. Finally the panel will conduct a hearing to obtain additional information concerning issues identified by the panel in the course of interviewing various client groups and studying various documents.

Organizations Providing Testimony

State of Michigan (Governor's Office)
State Board of Education
State Department of Education
Michigan Education Association
National Education Association
Detroit Federation of Teachers
Detroit Urban League
Council of State Organizations
League of Women Voters of Michigan
Michigan Art Education Association
Michigan Association of Elementary School Principals
Michigan Association for Health, Physical Education and Recreation
Michigan Association of Secondary School Principals
Michigan Association of School Administrators
Michigan Association for Supervision and Curriculum Development
Michigan Council of Teachers of Mathematics
Michigan Federation of Teachers
Michigan Science Teachers Association
Michigan Social Studies Council

Also teachers, administrators, and concerned citizens.